

علیرضا آخوندی

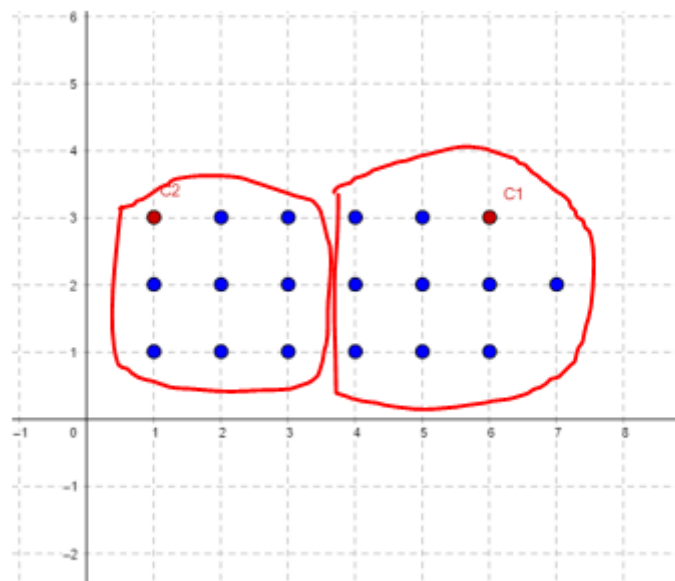
9731107

سوال اول :

فرمول فاصله L1 به صورت زیر است:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

در ابتدا خوشه ها به شکل زیر هستند:



C2:

$$X = \frac{1 \times 3 + 2 \times 3 + 3 \times 3}{9} = 2$$

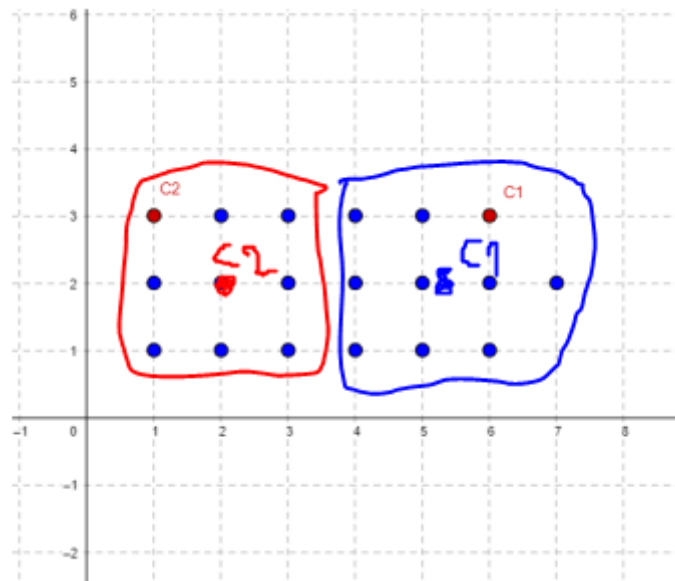
$$Y = \frac{1 \times 3 + 2 \times 3 + 3 \times 3}{9} = 2$$

C1:

$$X = \frac{4 \times 3 + 5 \times 3 + 6 \times 3 + 7 \times 1}{10} = 5.2$$

$$Y = \frac{1 \times 3 + 2 \times 4 + 3 \times 3}{10} = 2$$

حال خوشه ها به شکل زیر تبدیل می شوند:



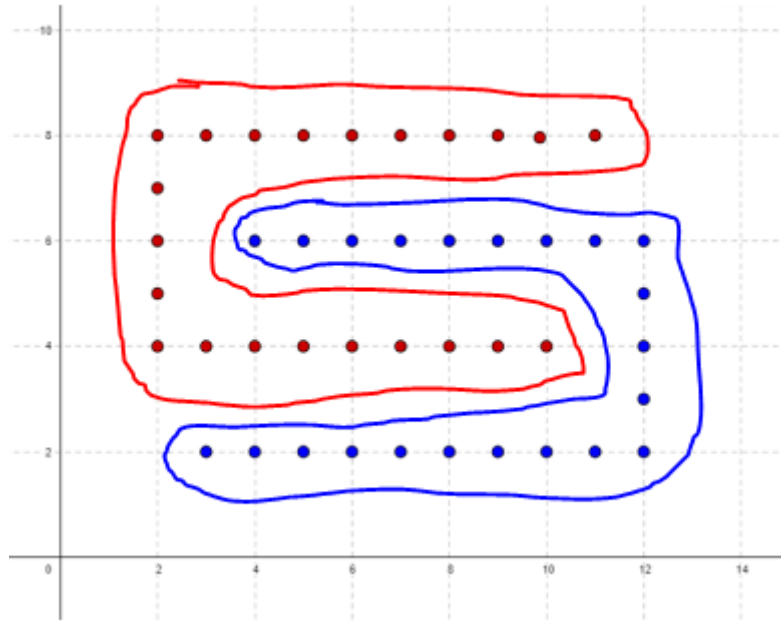
از آن جایی که شکل خوشه ها تغییری نکرده است به اجرا ادامه نمی دهیم.

سوال دوم: برای دسته بندی داده های مورد نظر از الگوریتم DBSCAN می توانیم استفاده کنیم. دلیل استفاده از این الگوریتم این است که مبنای خوشه بندی آن چگالی داده در یک ناحیه است و می تواند خوشه های پیچیده تری را نسبت به الگوریتم kmeans ایجاد کند و نسبت به outlier ها بهتر عمل می کند. حال اگر بخواهیم به خوشه بندی مشخص شده در شکل برسیم باید از این الگوریتم استفاده کنیم تا نواحی چگال را برای ما شناسایی کند.

حال برای خوشه بندی این مجموعه داده از پارامتر های زیر استفاده می کنیم:

$$\text{minPoints} = 2, \text{eps} = 1$$

پس از اجرای الگوریتم خروجی زیر حاصل می شود:



سوال سوم :

Closed Frequent Itemset ها زیرمجموعه frequent itemset ها محسوب می شوند. از طرفی این مجموعه، بسته بوده و مقدار support آن بزرگتر و یا مساوی minSup است. همچنین itemset ای بسته تعریف می شود که هیچ superset ای از آن مجموعه وجود نداشته باشد که مقدار support یکسانی با آن داشته باشد.

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

بسته بودن X به صورت زیر تعریف می شود :

$$S(X) > S(Y)$$

حال اگر X یک closed frequent itemset باشد، می توان به یکی از نتایج زیر رسید:

$$S(X) \geq \text{minsup} > S(Y)$$

$$S(X) > S(Y) \geq \text{minsup}$$

اگر حالت اول رخ بدهد، که اصلاً Y frequent itemset نمی باشد. اگر حالت دوم رخ بدهد، Y frequent itemset می باشد. پس اگر closed frequent itemset ها و support آنها را داشته باشیم، می توانیم از رابطه دوم یعنی  $S(X) > S(Y) \geq \text{minsup}$  مقدار پشتیبانی frequent itemset ها را تعیین کنیم.

سوال چہارم:

سوال چارم.  $\text{minsup} = \frac{1}{3} \times 6 = 2$   
 $\text{set} = \{ \text{نارنگی، انار، دوز، دڙقال، دسيب} \}$

1- frequent

①	②	③	④	⑤
نارنگی	دوز	انار	دڙقال	دسيب
2	3	3	3	4
✓	✓	✓	✓	✓

2 frequent

انار، دڙقال	دوز، دڙقال	دسيب، نارنگی	دسيب، انار	دسيب، دوز	دسيب، دڙقال
1	2	2	1	2	2
1	2	2	1	2	2
نارنگی، انار	دوز، نارنگی	دوز، انار	نارنگی، دڙقال		
1	2	1	0		

$\text{set} = \{ \text{نارنگی، دڙقال، دوز، دسيب} \}$

3- frequent

نارنگی، دڙقال، دوز	نارنگی، دوز، دسيب	دڙقال، دوز، دسيب	دڙقال، دوز، دسيب
0	0	0	2
0	0	0	2

$$\text{سوال چهارم} \quad \checkmark \quad 0.6 > 0.66 \quad \frac{2}{3} = (\text{سیب} \rightarrow \text{پرتقال}) \quad \checkmark$$

$$\checkmark \quad (\text{پرتقال} \rightarrow \text{سیب}) = \frac{2}{4} = 0.5 \quad \times$$

$$\checkmark \quad (\text{موز} \rightarrow \text{سیب}) = \frac{2}{4} = 0.5 \quad \times$$

$$\checkmark \quad (\text{سیب} \rightarrow \text{موز}) = \frac{2}{3} = 0.66$$

$$\checkmark \quad (\text{سیب} \rightarrow \text{نارنگی}) = \frac{2}{4} = 0.5 \quad \times$$

$$\checkmark \quad (\text{سیب} \rightarrow \text{نارنگی}) = \frac{2}{2} = 1$$

$$\checkmark \quad (\text{پرتقال} \rightarrow \text{موز}) = \frac{2}{3} = 0.66$$

$$\checkmark \quad (\text{پرتقال} \rightarrow \text{موز}) = \frac{2}{3} = 0.66$$

$$\checkmark \quad (\text{موز} \rightarrow \text{سیب}) = \frac{2}{2} = 1$$

$$\checkmark \quad (\text{پرتقال} \rightarrow \text{سیب}) = \frac{2}{2} = 1$$

$$\checkmark \quad (\text{پرتقال} \rightarrow \text{موز}) = \frac{2}{2} = 1$$

(min)

	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>
P <sub>1</sub>	a	<del>0.4</del>	<del>0.6</del>		
P <sub>2</sub>	0.1	a			
P <sub>3</sub>	0.47	0.64	a		
P <sub>4</sub>	0.55	0.48	0.44	a	
P <sub>5</sub>	0.35	0.85	0.85	0.76	a

(1)

ادغام  
P<sub>1</sub>, P<sub>2</sub>

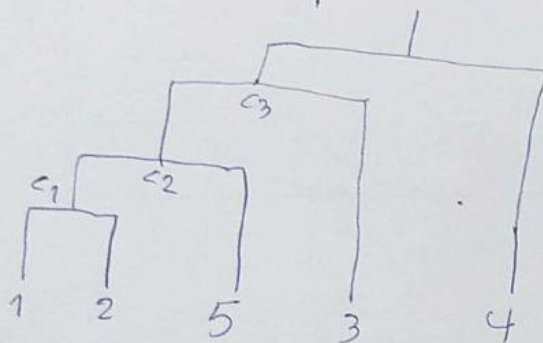
	C <sub>1</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>
C <sub>1</sub>	a			
P <sub>3</sub>	0.47	a		
P <sub>4</sub>	0.48	0.44	a	
P <sub>5</sub>	0.35	0.85	0.76	a

ادغام  
C<sub>1</sub> و P<sub>5</sub>

	C <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>
C <sub>2</sub>	a		
P <sub>3</sub>	0.47	a	
P <sub>4</sub>	0.48	0.44	a

ادغام  
C<sub>2</sub> و P<sub>3</sub>

	C <sub>3</sub>	P <sub>4</sub>
C <sub>3</sub>	a	
P <sub>4</sub>	0.44	a



(max)

ادغام  
P<sub>1</sub>, P<sub>2</sub>

	C <sub>1</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>
C <sub>1</sub>	a			
P <sub>3</sub>	0.64	a		
P <sub>4</sub>	0.55	0.44	a	
P <sub>5</sub>	0.98	0.85	0.76	a

ادغام  
P<sub>3</sub>, P<sub>4</sub>

	C <sub>1</sub>	C <sub>2</sub>	P <sub>5</sub>
C <sub>1</sub>	a		
C <sub>2</sub>	0.64	a	
P <sub>5</sub>	0.98	0.85	a

ادغام  
C<sub>2</sub> و C<sub>1</sub>

	C <sub>3</sub>	P <sub>5</sub>
C <sub>3</sub>	a	
P <sub>5</sub>	0.8	a

